Library Progress International *Print version ISSN 0970 1052*
Vol.44, No.4, Jul-Dec 2024: P. 737-747 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

AI-Based Modeling of Leaf Miner Incidence in Tomato Crops at Rajahmundry, India

Satish Kumar Yadav[1], D. Pawar[1], Latika Yadav[2], Anchal Yadav[2] Priyanka Mishra[3], Saurabh Tripathi[3]

[1]Department of Statistics, Amity Institute of Applied Sciences, Amity University, Noida-201313
[2] Vijay Singh Pathik Government Post Graduate College Kairana, Shamli Uttar Pradesh, India
[3]Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar, Uttarakhand

How to cite this article: Satish Kumar Yadav, D. Pawar, Latika Yadav, Anchal Yadav Priyanka Mishra, Saurabh Tripathi(2024). AI-Based Modeling of Leaf Miner Incidence in Tomato Crops at Rajahmundry, India. Library Progress International, 44(4), 737-747

Abstract
This study investigates the population dynamics of the leaf miner (*Liriomyza trifolii*) in tomato (*Solanum lycopersicum Linnaeus*) crops over eight consecutive years (2011–2018) during the Kharif season, with a focus on the relationship between pest population and various weather parameters. The weather variables examined include maximum and minimum temperature (MaxT and MinT), morning and evening relative humidity (RHM and RHE), sunshine hours (SS), wind velocity (Wind), total rainfall (RF) and the number of rainy days (RD). The findings reveal that the highest average population of leaf miners (1.3 larvae per plant) was observed in the protected experimental field during the 31st Standard Meteorological Week (SMW) of 2012. In contrast, the lowest population (0.1 larvae per plant) was recorded in the unprotected experimental field in 2016. Correlation analysis highlighted that wind velocity and rainy days (both current and lagged) exhibited both negative and positive influences, respectively, on leaf miner incidence. Additionally, minimum temperature and evening relative humidity negatively impacted leaf miner populations, while maximum temperature and rainy days (current and lagged) had a highly significant positive effect on pest growth. To develop predictive models for leaf miner incidence, the study applied various machine learning techniques, including support vector regression (SVR), random forest (RF), and traditional statistical models such as multiple linear regression (MLR), general regression neural network (GRNN), and feedforward neural network (FFNN). The performance of these models was compared based on root mean square error (RMSE) values. Among the models, the random forest (RF) model outperformed others by yielding the lowest RMSE values, indicating superior prediction accuracy. The Diebold-Mariano (D-M) test was further employed to assess the forecasting performance of the applied models, and the random forest model was found to provide the most accurate predictions of leaf miner incidence. The analysis was conducted using the R programming language. In conclusion, demonstrates that weather variables, particularly maximum temperature and rainy days, significantly affect leaf miner populations in tomato crops. The random forest model proved to be the most effective tool for predicting pest incidence, offering valuable insights for integrated pest management strategies in agriculture.
Keywords: Accuracy, Machine Learning, Leaf minar, Weather.

Introduction

Tomato (*Solanum lycopersicum Linnaeus*), native to South America, is one of the most economically significant and widely consumed crops globally. It is not only a key ingredient in culinary dishes but also a vital source of dietary antioxidants, such as lycopene, which is associated with numerous health benefits, including cancer prevention and improved heart health. Globally, tomato cultivation spans across 4.78 million hectares, yielding a production of 177.0 million tonnes and an average productivity of 37.0 tonnes per hectare (Anon, 2018). The major tomato-producing countries include China, India, the USA, Turkey, Egypt, Iran, Italy, and Spain. India, a key player in global tomato production, cultivates the crop throughout the year, across diverse agro-climatic regions. In India, tomato cultivation is spread over 0.79 million hectares, producing approximately 19.76 million tonnes annually (Anon, 2018). The primary tomato-producing states include Madhya Pradesh, Karnataka, Andhra Pradesh, Telangana, Odisha, Gujarat, and West Bengal. Telangana alone contributes significantly to this production, with tomatoes cultivated on 0.41 lakh hectares, yielding a production of 1.17 million tonnes at a productivity rate of 28.2 tonnes per hectare. Despite its significant contribution, tomato production in India faces several challenges, particularly related to pests and diseases, which drastically reduce yield.

Tomato Pest Infestation: A Major Constraint

Among the many challenges in tomato farming, pest infestation is one of the most critical constraints. Pests not only reduce crop yield but also affect the quality of the fruit, making it less marketable. One of the most destructive pests in recent years is the tomato pinworm, *Tuta absoluta*, which has caused severe damage to tomato crops in various regions of India. The tomato pinworm, an invasive pest originally from South America, first appeared in the Malnad and Hyderabad-Karnataka regions of Karnataka (Sridhar *et al.,* 2014). By November 2014, it had spread to Telangana, where it caused extensive crop losses of up to 60% (Kumari et al., 2015b). The pest feeds on tomato plants in a highly destructive manner, especially during its larval stages. The first two instars (larval stages) feed on the mesophyll (the inner tissue of the leaf) while leaving the epidermis (the outer layer) intact. This feeding pattern creates tunnels known as "mines" on the leaves. As the larvae progress to the third and fourth in stars, they become more aggressive, bored into stalks, apical buds, and fruits. This leads to significant damage to the plants and fruit, rendering the crop unmarketable. Tomatoes infested by *Tuta absoluta* are easily identifiable by the characteristic pinholes on the surface of the fruit. The destructive potential of this pest is staggering. In certain regions of Telangana, up to 90% of the tomato crop has been lost due to *Tuta absoluta* infestations (Kumari *et al.,* 2018). Effective management and control of this pest are therefore essential to safeguarding tomato production and minimizing economic losses.

The Role of Climatic Factors in Pest Incidence

The occurrence and development of insect pests such as the tomato pinworm are heavily influenced by climatic factors, including temperature, relative humidity, and precipitation (Aheer *et al.,* 1994). These factors play a critical role in determining the lifecycle, population dynamics, and distribution of pests. For example, higher temperatures can accelerate the development of pests, allowing them to reproduce and spread more rapidly. Similarly, changes in relative humidity and rainfall can affect pest survival rates, as well as their movement and distribution across different regions. Understanding the influence of climatic variables on pest incidence is crucial for developing effective pest management strategies. By analyzing historical weather data and correlating it with pest occurrence, predictive models can be developed to forecast future pest outbreaks. This enables farmers to implement timely and effective control measures, thereby reducing crop losses and increasing yield.

Machine Learning in Agriculture: A New Frontier

In recent years, advancements in data mining and machine learning technologies have revolutionized various fields, including agriculture. Machine learning is a form of artificial intelligence (*AI*) that enables computers to learn from data, identify patterns, and make predictions without being explicitly programmed for each task. This technology is increasingly being applied in agriculture to optimize farming practices, improve yield predictions, and develop more effective pest management strategies. Machine learning is particularly useful for analyzing complex datasets, such as those related to climatic factors and pest incidence. Traditional statistical models, such as multiple linear regression (MLR), are often limited by their reliance on linear relationships between variables. In contrast, machine learning algorithms can handle nonlinear relationships and interactions between multiple variables, making them more suitable for analyzing complex agricultural data (Paswan and Begum, 2013).

Predictive Modeling for Tomato *Tuta absoluta* Infestation

In this study, machine learning techniques were applied to forecast the incidence of *Tuta absoluta* using nine years of weather data (2011–2018). The goal was to develop a predictive model that could accurately forecast pest outbreaks based on climatic factors, enabling farmers to implement timely pest control measures. Several machine learning algorithms were tested, including: Support Vector Regression (SVR): SVR is a type of supervised learning algorithm used for regression analysis. It works by finding a hyperplane that best fits the data while minimizing the error between predicted and actual values. SVR is particularly effective for modeling nonlinear relationships. Generalized Regression Neural Network (GRNN): GRNN is a type of artificial neural network designed for regression problems. It can approximate any arbitrary function, making it a powerful tool for modeling complex systems such as pest dynamics. Random Forest (RF): Random Forest is an ensemble learning technique that constructs multiple decision trees and averages their outputs to improve prediction accuracy. It is particularly useful when dealing with large datasets and complex variable interactions.

Feedforward Neural Network (FFNN): FFNN is a type of neural network where information flows in one direction, from input to output. It is commonly used for pattern recognition and predictive modeling. Multiple Linear Regression (MLR): MLR is a traditional statistical method used to explore the relationship between a dependent variable (e.g., pest incidence) and two or more independent variables (e.g., temperature, humidity). While MLR is widely used in agricultural research, its performance may be limited when dealing with nonlinear relationships. The performance of the various machine learning models was evaluated using root mean square error (RMSE), a metric that measures the difference between predicted and observed values. Lower RMSE values indicate better model accuracy. Among the models tested, Random Forest (RF) consistently outperformed the other algorithms, producing the lowest RMSE values. This indicates that RF was the most accurate predictor of *Tuta absoluta* incidence. The superior performance of RF can be attributed to its ability to handle large datasets and complex interactions between climatic variables. By constructing multiple decision trees and averaging their outputs, RF reduces the likelihood of overfitting and improves the stability of its predictions. In contrast, MLR showed limited accuracy in forecasting pest incidence, likely due to its inability to capture the nonlinear relationships between pest occurrence and climatic factors. While SVR and GRNN performed reasonably well, they did not match the accuracy of RF. To further validate the model performance, the Diebold-Mariano (D-M) test was applied to compare the forecasting accuracy of the various models. The D-M test confirmed that Random Forest had the highest predictive accuracy, making it the most reliable model for forecasting *Tuta absoluta* outbreaks. This study highlights the potential of machine learning techniques, particularly Random Forest, in predicting the incidence of tomato pinworm (*Tuta absoluta*) based on climatic factors. Accurate and timely predictions are essential for farmers to implement effective pest management strategies, minimize crop losses, and optimize tomato yields. The integration of machine learning in agriculture represents a significant advancement in precision farming. By leveraging large datasets and advanced algorithms, farmers can make data-driven decisions that enhance productivity and sustainability. As climatic conditions continue to change, the importance of predictive modeling will only increase, helping farmers adapt to new challenges and safeguard their crops against pests like *Tuta absoluta*.

MATERIALS AND METHODS

Multiple Linear Regression Analysis

The general form of MLR for a data set of N observations on a response variable Y and p predictor variables, $X_1, X_2, \dots, X_p$ is

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon.$$

where β_0 is intercept, $\beta_1, \dots \beta_p$ are the regression coefficients and ε is error term which is assumed to follow the normal distribution with mean zero and a constant variance. In the present investigation, stepwise selection procedure for selecting the significant variable in the model was adopted.

Support vector regression (SVR)

For a given data set $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in R^n$ input vector is, $y_i \in R$ is scalar output and N corresponds to size of data set, general form of Nonlinear SVR estimating function (Fig. 1) is:

$$f(x) = w^T \varphi(x) + b$$

Where $\varphi(.): R^n \to R^{n_h}$ is a nonlinear mapping function from original input space into a higher dimensional feature space, which can be infinitely dimensional, $w \in R^{n_h}$ is weight vector, b is bias term and superscript T indicates transpose.

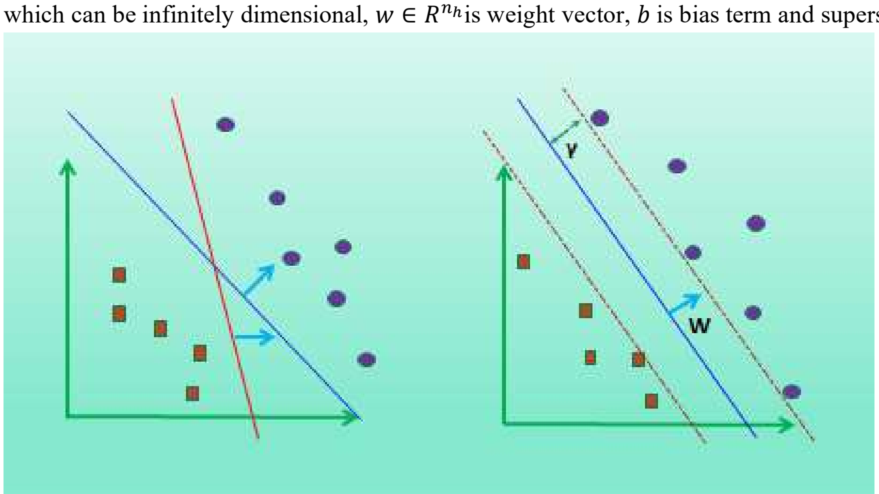

Figure 1: A schematic representation of Vapnik ε-insensitive loss function and accuracy tube under non-linear SVR model set-up

The coefficients w and b are estimated from data by minimizing the following regularized risk function:

$$R(\theta) = \frac{1}{2}\left||w|\right|^2 + C\left[\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)\right]$$

In above equation, first term $\frac{1}{2}||w||^2$ is called 'regularised term', which measures flatness of the function. Second term $\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)$ called 'empirical error' is estimated by Vapnik ε-insensitive loss function, C referred to as regularized constant. The Support Vector Regression (SVR) model was applied using the R software package (e1071) to analyze the severity of early blight in tomato crops. This model utilized data on mean and maximum severity for available seasons at each location, incorporating weather variables lagged by one and two weeks. These weather variables, which included factors such as temperature and humidity, were also considered in the Multiple Linear Regression (MLR) analysis. By employing SVR, the study aimed to enhance the accuracy of predictions regarding early blight severity, ultimately assisting farmers in implementing timely and effective disease management strategies.

Artificial neural network (ANN)

Artificial Neural Networks (ANN) are powerful tools for modeling complex, nonlinear relationships, making them highly effective when the underlying data patterns are unknown. ANN are self-adaptive, which means they can adjust to the data as they learn through training. A typical ANN consists of three main components: the input layer, which receives external data; one or more hidden layers, where computation and pattern recognition occur; and the output layer, which delivers the final predicted value. Each layer is composed of nodes (or neurons), and, apart from the input nodes, every node is a neuron that applies a nonlinear activation function to transform inputs and capture intricate patterns. The most widely used type of ANN is the multi-layer perceptron (MLP), a class of feedforward neural networks where information moves in one direction, from input to output, without looping back. MLP are especially effective for supervised learning tasks such as classification and regression. In MLP, neurons in each layer are connected to the neurons in the next layer, allowing for the learning of complex, nonlinear mappings between input and output. The backpropagation algorithm is typically used for training MLPs, where the model adjusts its weights based on the error between the predicted and actual outputs, iterating until the error is minimized. One of the key strengths of ANN and MLP is their ability to approximate any continuous function given sufficient data and complexity. This makes them versatile tools for a wide range of applications, from image and speech recognition to financial forecasting and climate modeling. In agriculture, ANN has been employed for predicting crop yields, classifying soil types, and forecasting pest outbreaks. For instance, Paul and Sinha (2016) applied ANN models to agricultural data, demonstrating their effectiveness in capturing nonlinear relationships and improving prediction accuracy. The flexibility and adaptability of ANN make them particularly well-suited for dynamic environments, where patterns are constantly changing, and traditional linear models may fail to capture the underlying complexity. Despite their effectiveness, ANN requires large datasets for training, and their "black box" nature can make them difficult to interpret compared to simpler, more transparent models. Nonetheless, their ability to model complex systems has made them indispensable in modern data science.

Random forest (RF)

Random Forest is a highly flexible and user-friendly machine learning algorithm that consistently delivers impressive results, often without the need for extensive hyperparameter tuning. Its simplicity and versatility make it one of the most widely utilized algorithms in the field of machine learning. As a supervised learning algorithm, Random Forest constructs multiple decision trees during training and merges their outputs to generate more accurate and stable predictions. This ensemble method helps mitigate the risk of overfitting, a common issue with individual decision trees, thereby enhancing the model's generalization to new, unseen data. One of the significant advantages of Random Forest is its applicability to both classification and regression problems, which constitute the majority of current machine learning applications. In classification tasks, Random Forest can effectively handle binary or multi-class problems, allowing it to categorize data points based on their features. For example, it might be used in healthcare to predict whether a patient has a specific disease based on their medical history and test results. Conversely, in regression tasks, Random Forest predicts continuous outcomes, such as estimating house prices based on various attributes like size and location. This dual capability makes it a valuable tool for practitioners across diverse domains, from finance and marketing to agriculture and environmental science. Another reason for the algorithm's popularity is its robustness against noise and outliers in datasets. The reliance on multiple decision trees means that the influence of any single data point is minimized, allowing the model to maintain high accuracy even when the dataset contains anomalies. Random Forest also includes built-in methods for assessing feature importance, helping users identify which variables significantly contribute to the model's predictions. This feature enhances interpretability and aids practitioners in understanding the factors driving their results. Despite its advantages, Random Forest has some limitations. While it is generally more interpretable than complex models like deep neural networks, it can still be considered a "black box," making it challenging to discern the exact decision-making process behind individual predictions. Additionally, the algorithm can be computationally intensive, particularly with large datasets or many trees, posing challenges regarding processing time and resource usage. In summary, Random Forest is a powerful and versatile machine learning algorithm that excels in both classification and regression tasks. Its ease of use, robustness to noise, and ability to provide insights into feature importance contribute to its widespread application across numerous fields. As machine learning continues to evolve, Random Forest remains a foundational tool for practitioners aiming to develop accurate predictive models while balancing complexity and interpretability.

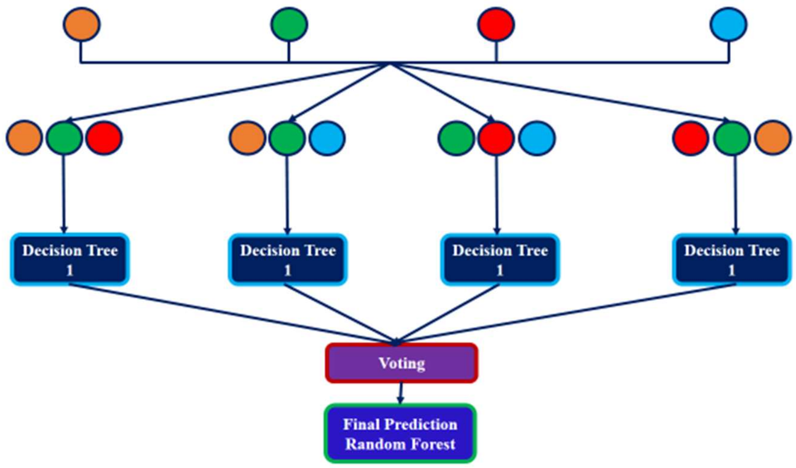

Figure 2: Workflow of random forest regression machine learning algorithm

Generalized regression neural network (GRNN)

Generalized Regression Neural Network (GRNN) is a type of neural network closely related to radial basis function networks and is fundamentally based on kernel regression principles. It can be viewed as a normalized radial basis neural network where each hidden neuron is centered around a training case. In GRNN, the radial basis function units typically represent probability density functions, with Gaussian functions being the most common choice (Celikoglu, 2006). This design allows GRNN to effectively approximate any arbitrary function that relates input vectors to target outputs, making it particularly powerful for regression tasks. One of the key advantages of GRNN is its ability to train rapidly and converge towards the optimal regression surface as the volume of training data increases (Specht, 1991). This feature is crucial in practical applications where large datasets are prevalent, as it enables GRNN to deliver accurate predictions without extensive computational requirements. The architecture of a GRNN consists of four distinct layers: an input layer that receives the input features, a hidden layer where the radial basis functions operate, a summation layer that aggregates the outputs from the hidden layer, and finally, an output layer that produces the final prediction. This multi-layer structure facilitates the smooth transition of data from inputs to outputs, enhancing the network's predictive capabilities. The GRNN's architecture not only contributes to its fast learning and adaptability but also allows it to maintain high levels of accuracy across various applications, including time series forecasting, financial prediction, and environmental modeling. Overall, the combination of rapid training, effective function approximation, and a straightforward architecture makes GRNN an advantageous tool for performing predictions in numerous domains.

Feed forward neural networks (FFNN)

Deep feedforward networks, commonly known as multilayer perceptron's (MLPs), serve as the foundation for most deep learning models and are widely utilized in various machine learning applications. These networks are primarily designed for supervised learning tasks, where the target function—i.e., the desired output that the network aims to predict—is already known. MLPs consist of multiple layers, including an input layer, one or more hidden layers, and an output layer, which allows them to learn complex mappings from input features to target outcomes. Each neuron in these layers applies a nonlinear activation function, enabling the network to model intricate relationships in the data. The architecture of deep feedforward networks facilitates the hierarchical learning of features, where lower layers capture simple patterns and higher layers learn more abstract representations. This capability is particularly valuable for tasks such as image recognition, natural language processing, and speech recognition, where the relationships between inputs and outputs can be highly complex and nonlinear. By leveraging large datasets, MLPs can generalize well to unseen data, making them a powerful tool for predictive modeling. In practice, training deep feedforward networks involves the use of backpropagation, a technique that updates the weights of the connections between neurons based on the error in predictions. This iterative optimization process enables the network to minimize the loss function, ultimately improving its performance. The flexibility and effectiveness of MLPs in approximating functions make them essential for practitioners in the field of machine learning. As a result, they have become a cornerstone of many advanced deep learning architectures, paving the way for innovations across various domains and applications. Overall, deep feedforward networks play a crucial role in advancing our understanding of machine learning and its potential applications in solving real-world problems.

Validation of forecasts

The dataset comprising pest population and weather data was split into two parts for analysis at each location, allocating 90% of the observations for estimation (model development) and the remaining 10% for validation. This division ensures that the models are trained on a substantial portion of the data while retaining a separate set for testing their predictive accuracy. A comparative assessment of the prediction performance of various models including Multiple Linear Regression (MLR), Random Forest (RF), Generalized Regression Neural Network (GRNN), Feedforward Neural Network (FNN), and Support Vector Regression (SVR) was conducted. The models were evaluated based on their root mean square error (RMSE), a widely used metric that quantifies the difference between predicted and observed values. The RMSE values provide insight into the accuracy of each model, allowing for a thorough comparison of their performance in predicting pest populations based on the weather variables. The formulae used for calculating RMSE are illustrated in Figure 3, emphasizing the systematic approach taken to assess the effectiveness of the different modeling techniques employed in this study. This rigorous evaluation helps identify the most suitable model for predicting pest population dynamics in relation to weather conditions.

$$\text{RMSE}=\sqrt{1/h\sum_{i=1}^{h}\left\{y_{t+i}-\hat{y}_{t+i}\right\}^{2}}$$

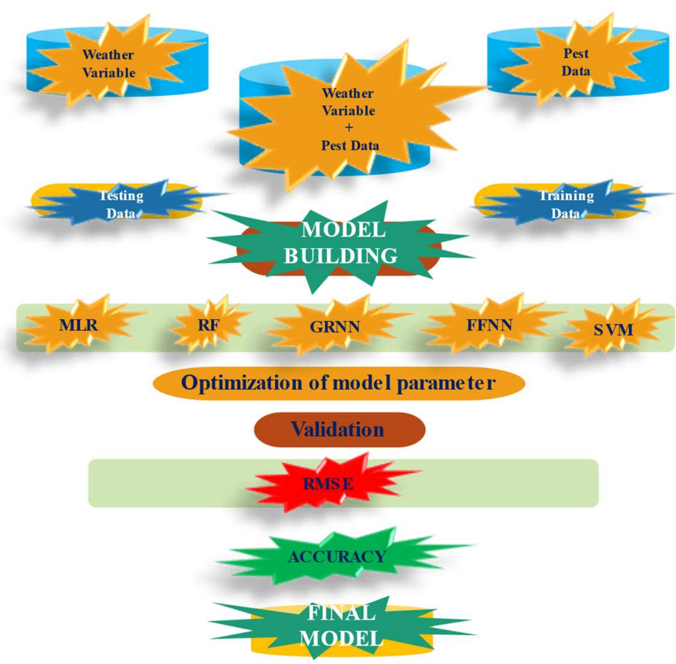

Figure 3: The above diagram showed the protocol followed in this study in implementation of machine learning.

where h denotes the number of observations for validation, yi is the observed value and $\hat{y}_i$ is the predicted one. The Diebold-Mariano test (Diebold and Mariano, 1995) was conducted for various pairs of models to assess differences in predictive accuracy between competing models, providing a statistical framework to evaluate which model performs better in terms of forecasting precision and reliability in predicting pest populations based on weather data.

Results and discussion

Seasonal dynamics and status of Leaf miner

In recent years, pest epidemics have increasingly been linked to climate change, highlighting the urgent need to understand how these environmental shifts affect crop pests in order to develop appropriate management strategies (Chowdappa, 2010). One significant focus of this research is the dynamics of leaf miner infestation in tomato crops, which was studied over eight consecutive years during the kharif season from 2011 to 2018 in Rajahmundry (Fig. 4). This study revealed that leaf miner infestations began as early as the 27th Standard Meteorological Week (SMW) in 2014, 2015, and 2016, while the latest initiation was observed in the 31st SMW during 2018. These findings indicate a varied incidence initiation period across the eight years of the study, suggesting that climate variability may influence the timing of pest outbreaks. Moreover, the peak population of leaf miners also exhibited significant variation across different seasons. The highest recorded peak occurred in 2012, with an infestation rate of 1.3 leaf miners per five leaves per plant, closely followed by 2014, which saw a peak of 1.2. In contrast, the lowest peak infestation was recorded in 2015, with only 0.3 leaf miners per five leaves per plant. This discrepancy in peak populations underscores the complex relationship between environmental factors and pest dynamics, with varying climatic conditions likely playing a crucial role in shaping these patterns. Throughout the duration of the study, the influence of weather factors on the peak occurrence of leaf miners became increasingly evident. Specifically, the maximum populations ranged from 1.1 to 1.3 leaf miners per five leaves per plant during the 29th to 34th SMW from 2012 to 2014, highlighting a significant increase in pest numbers during this period. In contrast, during 2015, the leaf miner population remained almost static, fluctuating between 0.1 and 0.3 per five leaves per plant, suggesting that adverse weather conditions or other environmental stressors may have hindered population growth. These observations underline the importance of monitoring environmental conditions and their potential impact on pest dynamics. The variability in infestation initiation and peak populations emphasizes the need for adaptive pest management strategies that account for climate change and its effects on pest life cycles and behaviours. By understanding the factors that contribute to leaf miner outbreaks, agricultural practices can be tailored to mitigate damage, ensuring that tomato production remains sustainable in the face of changing climatic conditions. In conclusion, the study of leaf miner infestation dynamics in tomato crops over an eight-year period reveals critical insights into how climate change influences pest behaviour and population fluctuations. The early initiation of infestations, variations in peak populations, and the relationship between weather factors and pest dynamics all point to the necessity for comprehensive pest management strategies that are responsive to the impacts of climate change. As agricultural systems continue to face the challenges posed by a changing climate, this research provides a foundational understanding that can guide future efforts to manage pests effectively and sustainably.

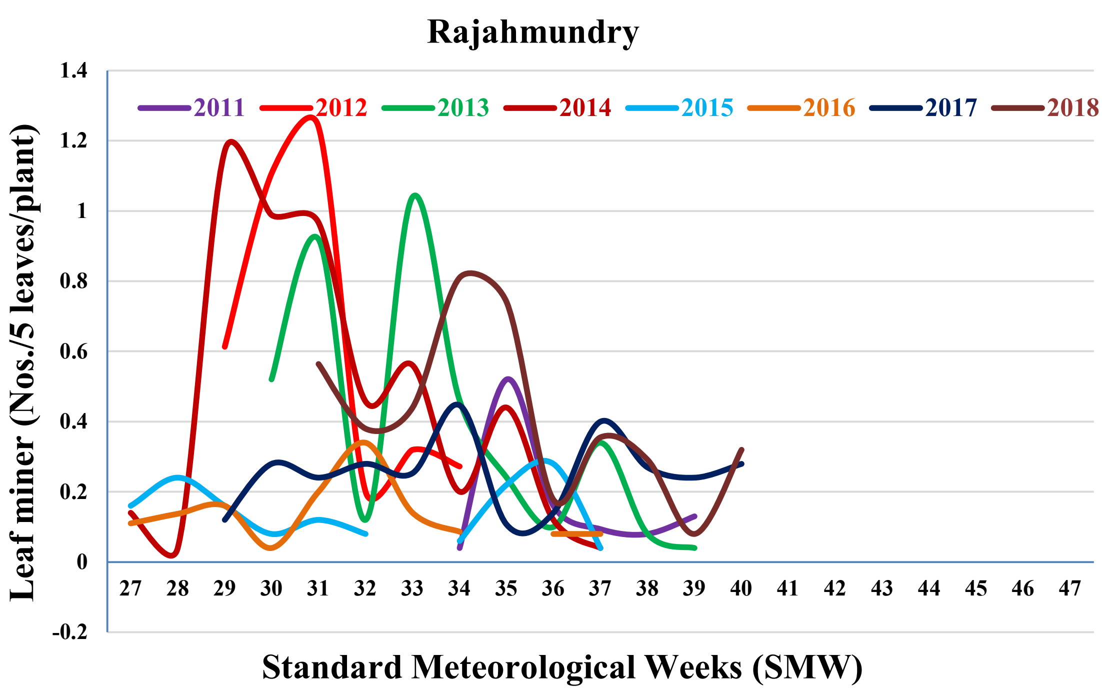

Figure. 4. The seasonal variation of occurrence of leaf miner in tomato

Comparative analysis of Leaf mineroccurrence across the years

The seasonal average serves as a crucial indicator of the severity of a pest in a particular locality. In this context, a comparative analysis of the mean incidence of leaf miner infestations across different seasons was conducted using Duncan's Multiple Range Test (DMRT), with the results summarized in Table 1. The findings reveal a notable variation in the seasonal averages, with the maximum leaf miner population reaching a significantly higher level of 0.62 during 2012. In contrast, the lowest incidences were recorded during 2015 and 2016, both at 0.14, indicating that these years experienced comparatively lower pest pressure. The results of the DMRT demonstrate the statistical significance of these differences, highlighting the dynamic nature of leaf miner populations in response to environmental conditions. This analysis not only underscores the importance of monitoring seasonal pest trends but also assists in developing effective management strategies tailored to specific years and conditions, thereby aiding farmers in mitigating the impact of leaf miners on tomato crops. By understanding these seasonal patterns, agricultural stakeholders can make informed decisions to enhance pest control efforts and improve overall crop yield and quality.

Table 1. Comparative analysis of Leaf miner occurrence across the years

2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
0.17^{c}	0.62^{a}	0.36^{abc}	0.47^{ab}	0.14^{c}	0.14^{c}	0.25^{bc}	0.39^{abc}

**** Means followed by the superscript of same at p<0.05 based on DMRT***

Correlation coefficients between leaf miner with weather factors

Pearson's correlation analysis was conducted to examine the influence of current and one-lag weather variables on the occurrence of leaf miners in tomato crops, as presented in Table 2. It is well established that insect pests and diseases are significantly governed by climatic and weather conditions, which directly impact their population dynamics. The analysis revealed that both current and one-lag wind speed had a negative influence on leaf miner incidence, while the number of rainy days (RainyD) exhibited a positive correlation. Additionally, the minimum temperature (MinT) and evening relative humidity (RHE) demonstrated a negative influence on the incidence of leaf miners, aligning with findings from a similar study by Choudary and Rosaiah (2000), which reported a negative correlation between minimum temperature and evening relative humidity with Liriomyza trifolii incidence in tomato. Notably, among all the weather variables analyzed, maximum temperature (MaxT) and current RainyD showed a highly significant positive effect on leaf miner populations. This finding contrasts with the results reported by Reddy and Kumar (2005), who found RainyD to have a different effect. However, the negative non-significant correlation obtained between morning and evening relative humidity aligns with Reddy and Kumar's (2005) observations, reinforcing the idea that humidity levels do not significantly impact leaf miner incidence. Overall, these correlations highlight the complex interplay between weather factors and pest populations, underscoring the importance of understanding these relationships for effective pest management strategies in tomato cultivation.

Table 2. Pearson Correlation Coefficients of Leaf miner occurrence with Climatic variables

Lag	**MaxT**	**MinT**	**RHM**	**RHE**	**Rainf**	**SunS**	**Wind**	**RainyD**
Current Lag	0.46***	-0.28*	0.10	-0.06	0.20	0.00	-0.32**	0.57***
One Lag	0.18	-0.19	0.14	-0.06	0.13	0.10	-0.32**	0.47***

****: significant at p< 0.01; *: significant at p< 0.05**

Validation

Once forecast values for the leaf miner population were obtained using five different models—namely Multiple Linear Regression (MLR), Random Forest (RF), Generalized Regression Neural Network (GRNN), Feedforward Neural Network (FFNN), and Support Vector Regression (SVR)—the performance of these predictions was assessed using the root mean square error (RMSE), as illustrated in Table 3. The results indicated that the RMSE values for the RF model were significantly lower than those of the other models, demonstrating its superior predictive accuracy. To ensure the adequacy of the fitted models, residual diagnostics were performed, revealing that there were no autocorrelations among the residuals, which supports the validity of the model assumptions. Furthermore, the population trend of leaf miners predicted by the RF model closely mirrored the actual observed values, as shown in Figure 5. This alignment between predicted and observed data highlights the effectiveness of the RF model in capturing the dynamics of leaf miner populations, thereby reinforcing its utility as a reliable tool for pest management strategies in agricultural practices. The findings suggest that employing advanced modeling techniques like RF can enhance the precision of pest forecasts, ultimately aiding farmers in making informed decisions regarding pest control measures.

Table 3. RMSE values of MLR, RF, GRNN, FNN and SVR models for predicting Leaf miner

Obs.	MLR	RF	GRNN	FFNN	SVR
0.30	0.28	0.35	0.08	0.10	0.30
0.54	0.47	0.45	0.52	0.40	0.41
1.13	0.64	0.80	0.09	0.07	0.46
0.24	0.50	0.54	0.14	0.13	0.28
0.47	0.32	0.15	0.24	0.34	0.21
0.41	0.14	0.30	0.24	0.16	0.33
0.10	0.06	0.24	0.08	0.06	0.32
0.40	0.29	0.12	0.17	0.23	0.16
0.44	0.12	0.28	0.44	0.26	0.28
0.48	0.11	0.16	0.12	0.18	0.11
RMSE	0.26	**0.23**	0.37	0.38	0.28

Obs.: Leaf miner, MLR: Multiple Linear Regression, SVR: Support vector regression, ANN: Artificial neural network, RF: Random Forest, GRNN: Generalized regression neural network and FFNN: Feed forward neural networks

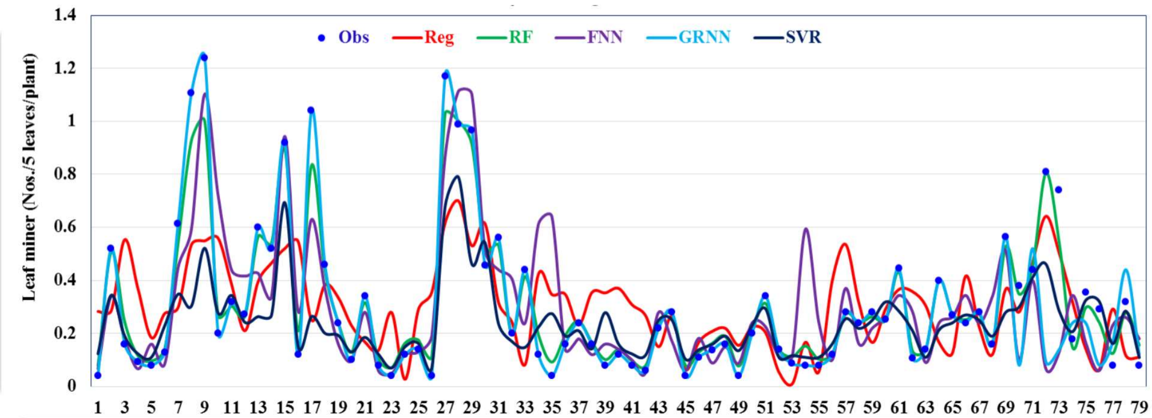

Figure 5 Plot of observed vs predicted by different models

Test results

The Diebold-Mariano test (Diebold and Mariano, 1995) was employed to compare the forecasting performance of the Random Forest (RF) model with Multiple Linear Regression (MLR), Support Vector Regression (SVR), Generalized Regression Neural Network (GRNN), and Feedforward Neural Network (FNN) models. This analysis was based on the null hypothesis that the predictive accuracy of the two competing models is equal. The results of various comparisons, including specific alternative hypotheses along with test statistics and their significance, are reported in Table 5. The findings revealed that the predictive accuracy of the MLR model was significantly lower than that of the RF model. Similar significant differences in predictive accuracy were observed for the other comparisons: RF vs. MLR, RF vs. SVR, RF vs. GRNN, and RF vs. FNN. These results imply that the RF model outperformed all other models in terms of predictive accuracy for the current dataset. Furthermore, the alternative hypotheses specified in Table 5 support this conclusion. Recent studies, such as that conducted by Balaban *et al.* (2019), have also shown that RF models can achieve over 99% accuracy in predicting the nymphal stage of the sun pest in the Middle Eastern region. This reinforces the growing evidence of the efficacy of RF in pest prediction and its potential utility in agricultural management strategies. Overall, the application of the Diebold-Mariano test in this context has provided valuable insights into the comparative effectiveness of different predictive models, highlighting the advantages of using RF for forecasting pest populations.

Table 4. Testing of Prediction Accuracy of Leaf miner

Combinations	Alternative Hypothesis	D-M Statistic	*p-value*
RF and MLR	Predictive accuracy of MLR is less than that of RF	-1.07	0.04
RF and SVR	Predictive accuracy of SVR is less than that of RF	-6.48	<0.0001
RF and GRNN	Predictive accuracy of GRNN is less than that of RF	-7.38	<0.0001
RF and FFNN	Predictive accuracy of FNN is less than that of RF	-1.08	0.03

*****: significant at p< 0.001; **: significant at p< 0.01; *: significant at p< 0.05**

Conclusion

Climate change significantly impacts the seasonal dynamics of pests and diseases, affecting agricultural productivity worldwide. In this context, the dynamics of leaf miner infestation in tomato crops were studied over eight consecutive kharif seasons from 2011 to 2018 in Rajahmundry. The present study revealed that leaf miner infestations appeared as early as the 27th standard meteorological week (SMW) in 2014, 2015, and 2016, with the latest appearance recorded by the 31st SMW in 2018. Moreover, the peak populations of leaf miners varied across different seasons, indicating the influence of environmental factors on pest dynamics. To analyze the leaf miner infestation patterns, statistical models, including Multiple Linear Regression (MLR) and advanced machine learning techniques such as Random Forest (RF), Generalized Regression Neural Network (GRNN), Feedforward Neural Network (FNN), and Support Vector Regression (SVR), were employed. Empirically, the RF model was found to be the most effective for predicting infestations within the present dataset, a conclusion that is further supported by the results of the Diebold-Mariano (DM) test. The methodologies discussed in this study have the potential to be replicated for forecasting the incidence of other significant pests and diseases affecting crucial agricultural crops. By implementing these predictive models, farmers can receive timely alerts and take necessary preventive measures to minimize losses due to pest and disease attacks, ultimately enhancing agricultural resilience in the face of climate change.

Reference

1. Aheer, G.M., Ahmad, K.J. and Ali, A. (1994). Role of weather in fluctuating aphid density in wheat crop. Ayub Agricultural Research Inst., Faisalabad (Pakistan).
2. Anon, Horticultural statistics at a glance (2018). Department of Agriculture, Cooperation & Farmers' Welfare Horticulture Statistics Division, p. 458.
3. Balaban, I., Acun, F., Arpalı, O.Y. Murat, F. Babaroglu, N.E., Akci, E., Çulcu, M., Ozkan, M. and Temizer, S. (2019). Development of a Forecasting and Warning System on the Ecological Life-Cycle of Sunn Pest.1905.01640.
4. Biau, G. (2012). Analysis of a Random Forests Model. *J. Mach. Learn. Res. 13*, 1063–1095.
5. Choudary, D.P. and Rosaiah, R.B. (2000). Seasonal occurrence of Liriomyza trifolii (Burgess) (Agromyzidae:Diptera) on tomato crop and its relation with weather parameters. *Pest Management Eco Zoology*. 8:91-95.
6. Chowdappa, P. (2010). Impact of climate change on fungal diseases of Horticultural crops: *In*: Challenges of climate change-Indian Horticulture (Singh, H. P, Singh, J.P, Lal, S. S. (eds.) Westville publishing house, New Delhi. p.144-151.
7. Fawagreh, F., Mohamed Medhat Gaber∗ and Eyad Elyan (2014). Systems Science & Control Engineering: An Open Access Journal, 2014 http://dx.doi.org/10.1080/21642583.2014.956265 Random forests: from early developments to recent advancements .
8. Haiguang Wang (2011). Prediction of Wheat Stripe Rust Based on Support Vector Machine, 2011 Seventh International Conference on Natural Computation, pp.378–382, 2011.
9. Kumari, D.A., Anitha, G., Anitha, V., Lakshmi, B.K.M., Vennila, S and Rao, N.H.P. (2015b). New record of leaf miner *Tuta absoluta* (Meyrich) in Tomato. *Insect Environment*. 20:136-138,.
10. Kumari, D.A., Anitha, G., Vennila, S. and Hanuman Nayak, M. (2018). Incidence of tomato pinworm, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in Telangana (India). *Journal of Entomology and Zoology Studies*; 6: 1085-1091.
11. Kumari, D.A., Lakshmi, B. ., Sireesha, K., Vennila, S. and Hariprasad, N., Rao. (2015a). Changing scenario of diseases in tomato in Telangana. 1st To 3rd 66 December, 2015 - Sri Konda Laxman Telangana State Horticultural University, Rajendranagar, Hyderabad, Telangana State.

12. Li, Y.H., Xu, J.Y., Tao, L., Li, X, F, Li, S. and Zeng, X. (2016). SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PloS one*;11: e0155290.
13. Paswan, R.P. and Begum, S.A. (2013). Regression and Neural Networks Models for Prediction of Crop Production. *Int. J. Sci. Eng. Res*. p 4- 11.
14. Reddy, N.A., and Kumar, C.T.A. (2005). Influence of weather factors on abundance and management of Serpentine leaf miner, Liriomyza trifolii (Bugress) on tomato. *Annals of Plant Protection Sciences*. 13:315-318.
15. Shekoofa, A., Emam, Y., Shekoufa, N., Ebrahimi, M. and Ebrahimie, E. (2014). Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: a new avenue in intelligent agriculture. *PloS one;* 9: e97288.
16. Sridhar, V., Chakravarthy, A.K., Asokan, R., Vinesh, L.S., Rebijith, K.B. and Vennila, S. (2014). New record of *Tuta absoluta* (Meyrick) (Lepidoptera*:* Gelechiidae) in India. *Pest Management in Horticultural Ecosystems*; 20:148-154.
17. Vennila, S., Satish Kumar Yadav, Priyanka Wahi, S., Kranthi, A., Amutha & Dharajothi, B. (2018). Seasonal Dynamics, Influence of Weather Factors and Forecasting of Cotton Sap Feeders in North India. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*; p467–476.
18. Younes C., (199). A generalized regression neural network and its application for leaf wetness prediction to forecast plant disease", Chemometrics and Intelligent Laboratory Systems, vol. 48, pp.47–58.